

# Physiological time-series analysis using approximate entropy and sample entropy

JOSHUA S. RICHMAN<sup>1,2</sup> AND J. RANDALL MOORMAN<sup>1</sup>

<sup>1</sup>Cardiovascular Division, Department of Internal Medicine, and Department of Molecular Physiology and Biological Physics, and Cardiovascular Research Center, University of Virginia Health Sciences Center, Charlottesville, Virginia 22908; and <sup>2</sup>Medical Automation Systems, Charlottesville, Virginia 22903

**Richman, Joshua S., and J. Randall Moorman.** Physiological time-series analysis using approximate and sample entropy. *Am J Physiol Heart Circ Physiol* 278: H2039–H2049, 2000.—Entropy, as it relates to dynamical systems, is the rate of information production. Methods for estimation of the entropy of a system represented by a time series are not, however, well suited to analysis of the short and noisy data sets encountered in cardiovascular and other biological studies. Pincus introduced approximate entropy (ApEn), a set of measures of system complexity closely related to entropy, which is easily applied to clinical cardiovascular and other time series. ApEn statistics, however, lead to inconsistent results. We have developed a new and related complexity measure, sample entropy (SampEn), and have compared ApEn and SampEn by using them to analyze sets of random numbers with known probabilistic character. We have also evaluated cross-ApEn and cross-SampEn, which use cardiovascular data sets to measure the similarity of two distinct time series. SampEn agreed with theory much more closely than ApEn over a broad range of conditions. The improved accuracy of SampEn statistics should make them useful in the study of experimental clinical cardiovascular and other biological time series.

probability; nonlinear dynamics

NONLINEAR DYNAMICAL ANALYSIS is a powerful approach to understanding biological systems. The calculations, however, usually require very long data sets that can be difficult or impossible to obtain. Pincus (21, 22) devised the theory and method for a measure of regularity closely related to the Kolmogorov entropy, the rate of generation of new information, that can be applied to the typically short and noisy time series of clinical data. This family of statistics, named approximate entropy (ApEn), is rooted in the work of Grassberger and Procaccia (10) and Eckmann and Ruelle (5) and has been widely applied in clinical cardiovascular studies (3, 6–8, 12–19, 24, 26, 29, 32–34).

The method examines time series for similar epochs: more frequent and more similar epochs lead to lower values of ApEn. Informally, given  $N$  points, the family of statistics  $\text{ApEn}(m, r, N)$  is approximately equal to the negative average natural logarithm of the conditional probability that two sequences that are similar for  $m$  points remain similar, that is, within a tolerance  $r$ , at the next point. Thus a low value of ApEn reflects a high degree of regularity. Importantly, the ApEn algo-

rithm counts each sequence as matching itself, a practice carried over from the work of Eckmann and Ruelle (5) to avoid the occurrence of  $\ln(0)$  in the calculations. This step has led to discussion of the bias of ApEn (22, 23, 27). In practice, we find that this bias causes ApEn to lack two important expected properties. First, ApEn is heavily dependent on the record length and is uniformly lower than expected for short records. Second, it lacks relative consistency. That is, if ApEn of one data set is higher than that of another, it should, but does not, remain higher for all conditions tested (22). This shortcoming is particularly important, because ApEn has been repeatedly recommended as a relative measure for comparing data sets (22–24).

To reduce this bias, we have developed and characterized a new family of statistics, sample entropy (SampEn), that does not count self-matches. SampEn is derived from approaches developed by Grassberger and co-workers (2, 9–11).  $\text{SampEn}(m, r, N)$  is precisely the negative natural logarithm of the conditional probability that two sequences similar for  $m$  points remain similar at the next point, where self-matches are not included in calculating the probability. Thus a lower value of SampEn also indicates more self-similarity in the time series. In addition to eliminating self-matches, the SampEn algorithm is simpler than the ApEn algorithm, requiring approximately one-half as much time to calculate. SampEn is largely independent of record length and displays relative consistency under circumstances where ApEn does not.

Cross-ApEn is a recently introduced technique for analyzing two related time series to measure the degree of their asynchrony (20, 28). Cross-ApEn is very similar to ApEn in design and intent, differing only in that it compares sequences from one series with those of the second. Because it does not compare a series with itself, bias from self-matches does not arise. A potential problem, however, remains in the necessity for each template to generate a defined, nonzero probability. Thus each template must find at least one match for  $m + 1$  points, or a probability must be assigned to it according to a “correction” strategy. We tested the effect of two extremes of correction strategies on cross-ApEn analysis. We find that cross-ApEn analysis lacks relative consistency, and conclusions about relative synchrony of pairs of time series depend on the unguided selection of analysis schemes. Cross-SampEn, on the

other hand, is defined as long as one template finds a match, and we find that cross-SampEn remains relatively consistent for conditions where cross-ApEn does not.

### THEORY

*ApEn reports on similarity in time series.* We employ the terminology and notation of Grassberger and Procaccia (10), Eckmann and Ruelle (5), and Pincus (21) in describing techniques for estimating the Kolmogorov entropy of a process represented by a time series and the related statistics ApEn and SampEn. The parameters  $N$ ,  $m$ , and  $r$  must be fixed for each calculation.  $N$  is the length of the time series,  $m$  is the length of sequences to be compared, and  $r$  is the tolerance for accepting matches. It is convenient to set the tolerance as  $r \times \text{SD}$ , the standard deviation of the data set, allowing measurements on data sets with different amplitudes to be compared. Throughout this work, all time series have been normalized to have  $\text{SD} = 1$ .

We proceed as follows: For a time series of  $N$  points,  $\{u(j): 1 \leq j \leq N\}$  forms the  $N - m + 1$  vectors  $\mathbf{x}_m(i)$  for  $\{i: 1 \leq i \leq N - m + 1\}$ , where  $\mathbf{x}_m(i) = \{u(i + k): 0 \leq k \leq m - 1\}$  is the vector of  $m$  data points from  $u(i)$  to  $u(i + m - 1)$ . The distance between two such vectors is defined to be  $d[\mathbf{x}(i), \mathbf{x}(j)] = \max\{|u(i + k) - u(j + k)|: 0 \leq k \leq m - 1\}$ , the maximum difference of their corresponding scalar components. Let  $B_i$  be the number of vectors  $\mathbf{x}_m(j)$  within  $r$  of  $\mathbf{x}_m(i)$  and let  $A_i$  be the number of vectors  $\mathbf{x}_{m+1}(j)$  within  $r$  of  $\mathbf{x}_{m+1}(i)$ . Define the function  $C_i^m(r) = (B_i)/(N - m + 1)$ . In calculating  $C_i^m(r)$ , the vector  $\mathbf{x}_m(i)$  is called the template, and an instance where a vector  $\mathbf{x}_m(j)$  is within  $r$  of it is called a template match.  $C_i^m(r)$  is the probability that any vector  $\mathbf{x}_m(j)$  is within  $r$  of  $\mathbf{x}_m(i)$ . The function  $\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln[C_i^m(r)]$  is the average of the natural logarithms of the functions  $C_i^m(r)$ . Eckmann and Ruelle (5) suggest approximating the entropy of the underlying process as  $\lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)]$ . Because of these limits, this definition is not suited to the analysis of the finite and noisy time series derived from experiments.

Pincus (21) saw that the calculation of  $\Phi^m(r) - \Phi^{m+1}(r)$  for fixed parameters  $m$ ,  $r$ , and  $N$  had intrinsic interest as a measure of regularity and complexity. He defines the related parameter  $\text{ApEn}(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)]$ , which for finite data sets is estimated by the statistic  $\text{ApEn}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r)$ . Algebraic manipulation reveals that  $\text{ApEn}(m, r, N) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln[C_i^m(r)] - (N - m)^{-1} \sum_{i=1}^{N-m} \ln[C_i^{m+1}(r)]$ . When  $N$  is large,  $\text{ApEn}(m, r, N)$  is approximately equal to  $(N - m)^{-1} \sum_{i=1}^{N-m} [-\ln(A_i/B_i)]$ , the average over  $i$  of the negative natural logarithm of the conditional probability that  $d[\mathbf{x}_{m+1}(i), \mathbf{x}_{m+1}(j)] \leq r$  given that  $d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r$ . The difference between  $\text{ApEn}(m, r, N)$  and this average logarithmic probability is less than  $(N - m + 1)^{-1} \ln(N - m + 1)$ , which is  $< 0.05$  for  $N - m + 1 \geq 90$  and  $< 0.02$  for  $N - m + 1 \geq 283$ . They differ because there are  $N - m + 1$  templates of length  $m$ , but only  $N - m$  templates of length  $m + 1$  (27). Thus, although the quantity  $C_{N-m+1}^m(r)$  is defined,

$C_{N-m+1}^{m+1}(r)$  is not, simply because the vector  $\mathbf{u}_{m+1}(N - m + 1)$  does not exist. For practical purposes, ApEn can be thought of as the negative natural logarithm of the probability that sequences that are close for  $m$  points remain close for an additional point. Because conditional probabilities lie between 0 and 1, the parameter  $\text{ApEn}(m, r)$  is a positive number of infinite range. For finite  $N$ , however, the largest possible value of  $\text{ApEn}(m, r, N)$  is  $-\Phi^{m+1}(r) \leq -\ln[(N - m)^{-1}]$ , so  $\text{ApEn}(m, r, N) \leq \ln(N - m)$ .

*ApEn(m, r, N) is biased and suggests more similarity than is present.* It is important to note that ApEn takes a template-wise approach to calculating this average logarithmic probability, first calculating a probability for each template. The ApEn algorithm thus requires that each template contribute a defined, nonzero probability. This constraint is overcome by allowing each template to match itself. Formally, because  $d[\mathbf{x}_m(i), \mathbf{x}_m(i)] = 0 \leq r$ , the ApEn algorithm counts each template as matching itself, a practice we will refer to as self-matching. This ensures that the functions  $C_i^m(r)$  are  $> 0$  for all  $i$ , thereby avoiding the occurrence of  $\ln(0)$  in the calculation. Thus  $\text{ApEn}(m, r, N)$  is certain to be defined under all circumstances. Pincus and Goldberger (23, 27) note that, as a consequence of this practice,  $\text{ApEn}(m, r, N)$  is a biased statistic. Formally, a statistic is biased if its expected value is not equal to the parameter it estimates. ApEn statistics are biased, because the expected value of  $\text{ApEn}(m, r, N)$  is less than the parameter  $\text{ApEn}(m, r)$  (27).

To discuss the bias caused by including self-matches, let us redefine the conditional probability associated with the template  $\mathbf{x}_i(m)$  by letting  $B_i$  denote the number of vectors  $\mathbf{x}_m(j)$  with  $j \neq i$ , such that  $d[\mathbf{x}_{m+1}(i), \mathbf{x}_{m+1}(j)] \leq r$  by  $A_i$ . The ApEn algorithm thus assigns to the template  $\mathbf{x}_i(m)$  a biased conditional probability of  $(A_i + 1)/(B_i + 1)$ , which is always greater than the unbiased  $A_i/B_i$ . In the limit as  $N$  approaches infinity,  $A_i$  and  $B_i$  will generally be large, making the biased and unbiased probabilities asymptotically equivalent. Therefore, this bias is evident only for the analysis of finite data sets and is a characteristic of the statistic  $\text{ApEn}(m, r, N)$ , rather than the parameter  $\text{ApEn}(m, r)$ . For a finite  $N$ , however, the result is that  $\text{ApEn}(m, r, N)$  is biased toward lower values of ApEn and returns values below those predicted by theory.

The largest deviation occurs when a large proportion of templates have  $B_i = A_i = 0$ , since these templates are assigned a conditional probability of 1, corresponding to perfect order. Furthermore, the difference between the biased and unbiased conditional probabilities assigned to individual templates makes the calculation sensitive to record length in a way that depends on the conditional probability. Suppose that the unbiased conditional probability is known and denote it by CP. For a given  $\mathbf{u}_m(i)$  let  $B_i$  denote the number of template matches without counting self-matches. The original algorithm estimates CP as  $(1 + A_i)/(1 + B_i) = (1 + B_i \times \text{CP})/(1 + B_i)$ . The fractional error of this

relative to CP is  $\text{Err} = \{[(1 + B_i \times \text{CP}) / (1 + B_i)] - \text{CP}\} / \text{CP}$ . To find the value of  $B_i$  necessary to keep the fractional error below a threshold  $\text{Err}_{\max}$ , we isolate  $B_i$  from the inequality  $\text{Err} \leq \text{Err}_{\max}$ , yielding  $B_i \geq [1 - \text{CP}(\text{Err}_{\max} + 1)] / (\text{CP} \times \text{Err}_{\max})$ . For independent, identically distributed (iid) random numbers, obtaining  $B_i$  matches of length  $m$  requires a data set containing, on average,  $B_i / (\text{CP})^m$  templates. For iid random numbers and  $m = 2$ , estimating  $\text{CP} = 0.368$  [ $\text{ApEn}(m, r, N) = 1$ ] within  $\text{Err}_{\max} = 0.05$  requires  $B_i \geq 33$  and a data set of  $>240$  points. Estimating  $\text{CP} = 0.135$  [ $\text{ApEn}(m, r, N) = 2$ ] with similar resolution requires  $B_i \geq 127$  and  $>6,900$  points.

The most straightforward way to eliminate the bias would be to remove self-matching from the ApEn algorithm, leaving it otherwise unaltered. However, without the inclusion of self-matches, the ApEn algorithm is not defined unless  $C_i^{m+1}(r) > 0$  for every  $i$ . Removing self-matches would make ApEn statistics highly sensitive to outliers; if there were a single template that matched no other vector, ApEn could not be calculated because of the occurrence of  $\ln(0)$ . For the set of uniform random numbers shown in Fig. 1A, without self-matches,  $\text{ApEn}(1, r, N)$  would not have been defined for  $r \leq 0.63$  and  $N < 4,000$ ;  $\text{ApEn}(2, r, N)$  would have required  $r \geq 1$  for  $N < 4,000$ . Thus, for many practical applications, self-matches cannot simply be excluded when ApEn statistics are calculated.

*Can the bias be corrected?* It is suggested that this bias can be reduced with a family of estimators  $\epsilon$  by defining  $C_{i\epsilon}^m(r) = (N - m + 1)^{-1} \{ \epsilon + \text{number of } j \neq i \text{ such that } d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r \}$  (27). Then  $\Phi_\epsilon^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} C_{i\epsilon}^m(r)$  and  $\text{ApEn}_\epsilon(m, r, N) = \Phi_\epsilon^m(r) - \Phi_\epsilon^{m+1}(r)$ . For large  $N$ , this is very close to estimating the conditional probabilities by  $(\epsilon + A_i) / (\epsilon + B_i)$ . When  $\epsilon < 1$ , the error due to counting self-matches is reduced, but this algorithm would still

report a conditional probability of 1 in the event that only self-matches are encountered. Another strategy would be to define  $\text{ApEn}_{\epsilon_1, \epsilon_2}(m, r, N) = \Phi_{\epsilon_2}^m(r) - \Phi_{\epsilon_1}^{m+1}(r)$  where  $\epsilon_1$  and  $\epsilon_2$  are small and  $\epsilon_1 \leq \epsilon_2$ . This is similar to estimating the conditional probabilities by  $(\epsilon_1 + A_i) / (\epsilon_2 + B_i)$ . Thus a template reporting only self-matches would be assigned a probability of  $\epsilon_1 / \epsilon_2$ . This remains a biased estimate of ApEn, with bias toward  $-\ln(\epsilon_1 / \epsilon_2)$  when a large proportion of templates report only self-matches. If  $\epsilon_1 / \epsilon_2 = (N - m + 1)^{-1}$  and  $A_i = B_i = 0$ , the bias is toward the lowest observable probability. This is the opposite bias of the original formulation of ApEn, where the bias is toward the highest possible probability. Still another approach would be to use the estimators  $\epsilon_1$  and  $\epsilon_2$  but incorporate them into the calculation only when  $A_i$  or  $B_i = 0$ , respectively. This would estimate the probabilities as  $A_i / B_i$  and would simply redefine  $A_i = \epsilon_1$  when  $A_i = 0$  and  $B_i = \epsilon_2$  when  $B_i = 0$ .

These approaches reduce the bias in the estimation of the individual conditional probabilities and ensure that perfect order would not be reported where none had been detected. None of the corrections eliminates bias; the bias is minimized only if  $\epsilon_1 / \epsilon_2 = \text{CP}$ , but CP is not known beforehand. Thus no family of estimators  $\epsilon$  that minimizes bias can be chosen a priori.

*SampEn statistics have reduced bias.* We developed SampEn statistics to be free of the bias caused by self-matching. The name refers to the applicability to time series data sampled from a continuous process. In addition, the algorithm suggests ways to employ sample statistics to evaluate the results, as explained below.

There are two major differences between SampEn and ApEn statistics. First, SampEn does not count self-matches. We justified discounting self-matches on the grounds that entropy is conceived as a measure of the rate of information production (5), and in this

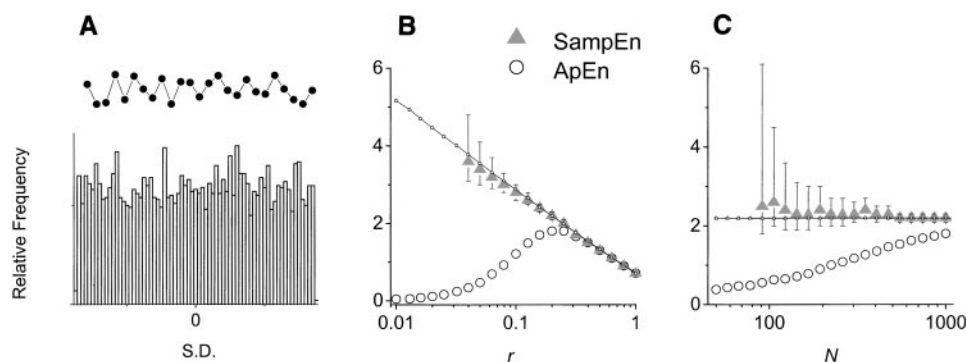


Fig. 1. Sample entropy (SampEn) and approximate entropy (ApEn) of random numbers with a uniform distribution. A: frequency histogram of 1,000 numbers. Inset: 25-point excerpt from time series. B: SampEn and ApEn as functions of  $r$  ( $m = 2$ ). Straight line, calculated theoretical values for ApEn and SampEn parameters. In this special case of uniformly distributed random numbers, their theoretical values coincide. Values of  $r$  are displayed logarithmically, so that theoretical values appear linear. C: SampEn and ApEn as functions of  $N$  ( $m = 2$ ,  $r = 0.2$ ). In B and C, confidence intervals for SampEn statistics are displayed as error bars. Straight line is theoretically predicted value of parameter SampEn(2,0.2). Sets of uniformly distributed random numbers were generated using a minimal standard number generator with added random shuffling (30), and all sets passed a runs test for random arrangement (1). For data shown above and for similar tests in which random data with different probabilistic distributions were used, theoretical values were calculated from expressions for SampEn( $m, r$ ) and ApEn( $m, r$ ) by use of trapezoidal numerical integration. We verified our calculation of ApEn statistics by comparing our results with published values for Henon and logistic map data (21).



context comparing data with themselves is meaningless. Furthermore, self-matches are explicitly dismissed in the later work of Grassberger and co-workers (2, 9, 11). Second, SampEn does not use a template-wise approach when estimating conditional probabilities. To be defined, SampEn requires only that one template find a match of length  $m + 1$ .

We began from the work of Grassberger and Procaccia (10), who defined  $C^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} C_i^m(r)$ , the average of the  $C_i^m(r)$  defined above. This differs from  $\Phi^m(r)$  only in that  $\Phi^m(r)$  is the average of the natural logarithms of the  $C_i^m(r)$ . They suggest approximating the Kolmogorov entropy of a process represented by a time series by  $\lim_{r \rightarrow 0} \lim_{N \rightarrow \infty} -\ln [C^{m+1}(r)/C^m(r)]$ ; self-matches are counted, and  $C^{m+1}(r)/C^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m} A_i/(N - m) \sum_{j=1}^{N-m+1} B_j$ .

In this form, however, the limits render it unsuitable for the analysis of finite time series with noise. We therefore made two alterations to adapt it to this purpose. First, we followed their later practice in calculating correlation integrals (2, 9, 11) and did not consider self-matches when computing  $C^m(r)$ . Second, we considered only the first  $N - m$  vectors of length  $m$ , ensuring that, for  $1 \leq i \leq N - m$ ,  $\mathbf{x}_m(i)$  and  $\mathbf{x}_{m+1}(i)$  were defined.

We defined  $B_i^m(r)$  as  $(N - m - 1)^{-1}$  times the number of vectors  $\mathbf{x}_m(j)$  within  $r$  of  $\mathbf{x}_m(i)$ , where  $j$  ranges from 1 to  $N - m$ , and  $j \neq i$  to exclude self-matches. We then defined  $B^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} B_i^m(r)$ . Similarly, we defined  $A_i^m(r)$  as  $(N - m - 1)^{-1}$  times the number of vectors  $\mathbf{x}_{m+1}(j)$  within  $r$  of  $\mathbf{x}_{m+1}(i)$ , where  $j$  ranges from 1 to  $N - m$  ( $j \neq i$ ), and set  $A^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} A_i^m(r)$ .  $B^m(r)$  is then the probability that two sequences will match for  $m$  points, whereas  $A^m(r)$  is the probability that two sequences will match for  $m + 1$  points. We then defined the parameter  $\text{SampEn}(m, r) = \lim_{N \rightarrow \infty} [-\ln [A^m(r)/B^m(r)]]$ , which is estimated by the statistic  $\text{SampEn}(m, r, N) = -\ln [A^m(r)/B^m(r)]$ . Where there is no confusion about the parameter  $r$  and the length  $m$  of the template vector, we set  $B = [(N - m - 1)(N - m)]/2$   $B^m(r)$  and  $A = [(N - m - 1)(N - m)]/2$   $A^m(r)$ , so that  $B$  is the total number of template matches of length  $m$  and  $A$  is the total number of forward matches of length  $m + 1$ . We note that  $A/B = [A^m(r)/B^m(r)]$ , so  $\text{SampEn}(m, r, N)$  can be expressed as  $-\ln (A/B)$ .

The quantity  $A/B$  is precisely the conditional probability that two sequences within a tolerance  $r$  for  $m$  points remain within  $r$  of each other at the next point. In contrast to  $\text{ApEn}(m, r, N)$ , which calculates probabilities in a template-wise fashion,  $\text{SampEn}(m, r, N)$  calculates the negative logarithm of a probability associated with the time series as a whole.  $\text{SampEn}(m, r, N)$  will be defined except when  $B = 0$ , in which case no regularity has been detected, or when  $A = 0$ , which corresponds to a conditional probability of 0 and an infinite value of  $\text{SampEn}(m, r, N)$ . The lowest nonzero conditional probability that this algorithm can report is  $2[(N - m - 1)(N - m)]^{-1}$ . Thus, the statistic  $\text{SampEn}(m, r, N)$  has  $\ln (N - m) + \ln (N - m - 1) - \ln (2)$

as an upper bound, nearly doubling  $\ln (N - m)$ , the dynamic range of  $\text{ApEn}(m, r, N)$ .

*Confidence intervals inform the implementation of SampEn statistics.*  $\text{SampEn}$  is not defined unless template and forward matches occur and is not necessarily reliable for small numbers of matches. We have reviewed  $\text{SampEn}(m, r, N)$  calculation as a process of sampling information about regularity in the time series and used sample statistics to inform us of the reliability of the calculated result. For example, say that we find  $B$  template matches, allowing for no more than  $B$  forward matches,  $A$  of which actually occur. We assign a value of 1 to the  $A$  forward matches and a value of 0 to the  $(B - A)$  potential forward matches that do not occur and compute the conditional probability measured by  $\text{SampEn}$  as the average of this sample of 0s and 1s. For operational purposes, we will assume that the sample averages follow a Student's  $t_d$  distribution, where  $d$  is the number of degrees of freedom. We can then say with 95% confidence that the "true" average conditional probability of the process is within  $\text{SD} t_{(B-1, 0.975)} / \sqrt{B}$  of the sample average, where  $\text{SD}$  is the sample standard deviation and  $t_{(B-1, 0.975)}$  is the upper 2.5th percentile of a  $t$  distribution with  $B - 1$  degrees of freedom (31). The size of the confidence intervals depends on the number  $B$  and the number of forward matches. Informally, large confidence intervals around  $\text{SampEn}(m, r, N)$  indicate that there are insufficient data to estimate the conditional probability with confidence for that choice of  $m$  and  $r$ . In addition, confidence intervals allow standard statistical tests of the significance of differences between data sets.

The confidence intervals for  $\text{SampEn}$  are displayed as error bars in the figures. For some small values of  $N$  and  $r$ , no value of  $\text{SampEn}$  is given. This indicates that  $B = 0$ ,  $A = 0$ , or the confidence intervals extended to a probability of  $>1$  or  $<0$ . In these cases, no value of  $\text{SampEn}(m, r, N)$  can be assigned with confidence.

*Cross-ApEn and cross-SampEn measure asynchrony.*  $\text{Cross-ApEn}$  is a recently introduced technique for comparing two different time series to assess their degree of asynchrony or dissimilarity (20, 28). The definition of  $\text{cross-ApEn}$  is very similar to  $\text{ApEn}$ . Given two time series of  $N$  points  $\{u(j): 1 \leq j \leq N\}$  and  $\{v(j): 1 \leq j \leq N\}$ , form the vectors  $\mathbf{x}_m(i) = \{u(i + k): 0 \leq k \leq m - 1\}$  and  $\mathbf{y}_m(i) = \{v(i + k): 0 \leq k \leq m - 1\}$ . The distance between two such vectors is defined as  $d[\mathbf{x}_m(i), \mathbf{y}_m(j)] = \max \{|u(i + k) - v(j + k)|: 0 \leq k \leq m - 1\}$ . Define  $C_i^m(r)(v||u)$  as the number of  $\mathbf{y}_m(j)$  within  $r$  of  $\mathbf{x}_m(i)$  divided by  $(N - m + 1)$ , then define  $\Phi^m(r)(v||u) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln [C_i^m(r)(v||u)]$ , and  $\text{cross-ApEn}(m, r, N)(v||u) = \Phi^m(r)(v||u) - \Phi^{m+1}(r)(v||u)$ . This is identical to the definition of the statistic  $\text{ApEn}$ , except the templates are chosen from the series  $u$  and compared with vectors from  $v$ . There is thus a directionality to this analysis, and we will call the series that contributes the templates the template series and the series with which they are compared the target series. We can then refer to  $\text{cross-ApEn}(m, r, N)$  (target||template).

We made two observations. First, because no template is compared with itself, there are no self-matches. Consequently,  $C_i^m(r)(v||u)$  can equal 0, and there is no assurance that cross-ApEn( $m, r, N$ )( $v||u$ ) will be defined. Second, there is a “direction dependence” of cross-ApEn analysis. We defined cross-SampEn to avoid both potential problems.

*Cross-ApEn is not always defined.* As noted above, cross-ApEn does not include self-matching and, thus, does not inherently suffer from the same bias as ApEn. A potential problem, however, remains in the necessity for each template to generate a defined, nonzero conditional probability. Thus each template must find at least one match for  $m + 1$  points, or a probability must be assigned to it. No guidelines have been suggested for handling this potential difficulty. Cross-SampEn, on the other hand, requires only that one pair of vectors in the two series match for  $m + 1$  points.

The family of MIX( $P$ ) stochastic processes (21) provided a testing ground for cross-ApEn. Informally, the MIX( $P$ ) time series of  $N$  points, where  $P$  is between 0 and 1, is a sine wave, where  $N \times P$  randomly chosen points have been replaced with random noise. We calculated cross-ApEn(1,  $r$ , 250) for the pair [MIX( $Q$ )||MIX( $P$ )] and its direction conjugate [MIX( $P$ )||MIX( $Q$ )] for 16 realizations of each of the 6 combinations of  $P = 0.1, 0.2, 0.3$  and  $Q = 0.5, 0.7$  over a range of values of  $r$  from 0.01 to 1.0. Cross-ApEn(1,  $r$ , 250) [MIX( $Q$ )||MIX( $P$ )] was not defined for any of the 96 pairs for  $r \leq 0.16$  and was defined for all of them only for  $r \geq 0.50$ . Cross-ApEn(1,  $r$ , 250) [MIX( $P$ )||MIX( $Q$ )] was not defined for any values of  $r \leq 0.32$  and was defined for all pairs only for  $r = 1.0$ .

To broaden the conditions for which cross-ApEn was defined, we introduced a correction factor into its algorithm. To avoid  $\ln(0)$  whenever  $C_i^m(r)(v||u) = 0$  or  $C_i^{m+1}(r)(v||u) = 0$ , we redefined them to be positive and nonzero. Rather than include these factors into the calculation of each  $C_i^m(r)(v||u)$  and  $C_i^{m+1}(r)(v||u)$ , we minimized their impact on the overall calculation by including them only when needed to ensure that cross-ApEn is defined. The cost of this adjustment, however, is the introduction of bias. For this reason, we tested cross-ApEn with two different correction strategies. The first strategy, which we called bias 0, was very similar to self-matching; we simply set any  $C_i^m(r)(v||u)$  or  $C_i^{m+1}(r)(v||u) = 1$  if it would otherwise have been 0. Thus, if a template matched no other at all, it would be assigned a conditional probability of 1, as with the original description of ApEn. If, however,  $C_i^m(r)(v||u) \neq 0$  but  $C_i^{m+1}(r)(v||u) = 0$ , we redefined  $C_i^{m+1}(r)(v||u) = (N - m)^{-1}$ , so that the probability assigned would be the lowest observable, nonzero probability given the nonzero value of  $C_i^m(r)(v||u)$ .

The second approach, which we called bias max, also only modified the functions  $C_i^m(r)(v||u)$  and  $C_i^{m+1}(r)(v||u)$  that would otherwise have been 0. Here,  $C_i^m(r)(v||u)$  was redefined to be 1 when it would otherwise have been 0 and  $C_i^{m+1}(r)(v||u)$  was redefined to be  $(N - m + 1)^{-1}$ , as for bias 0.

The only difference between the two strategies is that bias max assigns to a template yielding no matches at all a probability of  $(N - m)^{-1}$ , the lowest nonzero probability allowed by the length of the time series. Thus bias 0 sets the bias toward a cross-ApEn value of 0 in the absence of any matches, whereas bias max sets the bias toward the highest observable value of cross-ApEn.

*Cross-ApEn is direction dependent; cross-SampEn is not.* Because of the logarithms inside the summation,  $\Phi^m(r)(v||u)$  will not generally be equal to  $\Phi^m(r)(u||v)$ . Thus cross-ApEn( $m, r, N$ )( $v||u$ ) and its direction conjugate cross-ApEn( $m, r, N$ )( $u||v$ ) are unequal in most cases.

In defining cross-SampEn, we set  $B_i^m(r)(v||u)$  as  $(N - m)^{-1}$  times the number of vectors  $\mathbf{y}_m(j)$  within  $r$  of  $\mathbf{x}_m(i)$ , where  $j$  ranges from 1 to  $N - m$ . We then defined  $B^m(r)(v||u) = (N - m)^{-1} \sum_{i=1}^{N-m} B_i^m(r)(v||u)$ . Similarly, we set  $A_i^m(r)(v||u)$  as  $(N - m)^{-1}$  times the number of vectors  $\mathbf{y}_{m+1}(j)$  within  $r$  of  $\mathbf{x}_{m+1}(i)$ , where  $j$  ranges from 1 to  $N - m$ . We then defined  $A^m(r)(v||u) = (N - m)^{-1} \sum_{i=1}^{N-m} A_i^m(r)(v||u)$ . Finally, we set cross-SampEn( $m, r, N$ )( $v||u$ ) =  $-\ln [A^m(r)(v||u) / B^m(r)(v||u)]$ . Examining this definition for direction dependence, we see that  $(N - m) B_i^m(r)(v||u)$  is the number of vectors from  $v$  within  $r$  of the  $i$ th template of the series  $u$ . Summing over the templates, we see that  $\sum_{i=1}^{N-m} (N - m) B_i^m(r)(v||u)$  simply counts the number of pairs of vectors from the two series that match within  $r$ . The number of pairs that match is clearly independent of which series is the template and which is the target. Because the last summation is equal to  $(N - m)^2 B^m(r)(v||u)$ , it follows that  $B^m(r)(v||u)$  is also direction independent, implying that cross-SampEn( $m, r, N$ )( $v||u$ ) = cross-SampEn( $m, r, N$ )( $u||v$ ). We further noted that cross-SampEn will be defined provided that  $A^m(r)(v||u) \neq 0$ . Cross-SampEn, on the other hand, requires only that one pair of vectors in the two series match for  $m + 1$  points.

We calculated cross-SampEn(1,  $r$ , 250) for the same realizations of the [MIX( $Q$ )||MIX( $P$ )] pairs used above to test cross-ApEn and over the same range of  $r$ . In contrast to cross-ApEn(1,  $r$ , 250) [MIX( $P$ )||MIX( $Q$ )] and cross-ApEn(1,  $r$ , 250) [MIX( $Q$ )||MIX( $P$ )], cross-SampEn(1,  $r$ , 250) [MIX( $P$ )||MIX( $Q$ )], which is identical to cross-SampEn(1,  $r$ , 250) [MIX( $Q$ )||MIX( $P$ )], was defined for all 96 pairs over the entire range of  $r$  considered.

*ApEn and SampEn can be calculated analytically for series of random numbers.* ApEn and SampEn derive from formulas suggested to estimate the Kolmogorov entropy of a process represented by a time series. At their root, each is a measurement of the conditional probability that two vectors that are close to each other for  $m$  points will remain close at the next point. There are several models, including sets of iid random numbers, for which the theoretical values of the parameters ApEn( $m, r$ )(21) and SampEn( $m, r$ ) can be calculated. We show here the case of uniform random numbers, for which the theoretical values of ApEn and SampEn are nearly identical.

The expected value of the key probability can be calculated analytically for series of iid numbers based only on their probabilistic distribution. The numbers' independence implies that the probability that two randomly selected sequences within  $rSD$  of each other for the first  $m$  points will remain within  $r$  at their next points is simply the probability that any two points will be within a distance  $rSD$  of each other. For random numbers with density function  $p(x)$  and standard deviation  $SD$ , the expression for this probability is  $\int_{-\infty}^{+\infty} p(t) [\int_{t-rSD}^{t+rSD} p(x) dx] dt$ . Because the parameter  $\text{ApEn}(m, r)$  measures the average of the negative logarithm of this probability, it is expected to return a value of  $-\int_{-\infty}^{+\infty} p(t) \ln [\int_{t-rSD}^{t+rSD} p(x) dx] dt$  (21).  $\text{SampEn}(m, r)$ , on the other hand, measures the logarithm of the conditional probability and is thus expected to return a value of  $-\ln [\int_{-\infty}^{+\infty} p(t) [\int_{t-rSD}^{t+rSD} p(x) dx] dt]$ .

Because these expressions for the expected values of the parameters  $\text{ApEn}$  and  $\text{SampEn}$  depend solely on  $r$  and the probabilistic character of the data, uniform random numbers provide a benchmark for testing the estimation of  $\text{ApEn}$  over a range of parameters. In particular,  $\text{ApEn}$  and  $\text{SampEn}$  are expected to give identical results for uniformly distributed random numbers. Figure 1A shows a histogram of the random numbers used and an excerpt from the sequence (*inset*). Taking the derivative of  $-\int_{-\infty}^{+\infty} p(t) \ln [\int_{t-rSD}^{t+rSD} p(x) dx] dt$  with respect to  $\ln(r)$  reveals that, for small  $r$  (in this case  $r \leq 1$ ),  $\text{ApEn}(m, r)$  decreases in proportion to  $\ln(r)$  for small  $r$ . This result generalizes to random numbers with other distributions provided that their density functions are smoothly continuous and bounded. The expected values are shown as the straight line in Fig. 1B. For random numbers with other distributions, the theoretical values of  $\text{ApEn}$  and  $\text{SampEn}$  will differ slightly. This is explained by noting that because  $-\ln$  is a convex function, Jensen's inequality (4) applied to each of the integral expressions implies that the  $\text{ApEn}$  is expected to return a value greater than or equal to  $\text{SampEn}$ .

## RESULTS AND DISCUSSION

*SampEn agrees with theory more closely than ApEn.* For most processes,  $\text{ApEn}$  statistics are expected to have two properties. First, the conditional probability that sequences within  $r$  of each other remain within  $r$  should decrease as  $r$  decreases and the criterion for matching becomes more stringent. In other words,  $\text{ApEn}(m, r, N)$  should increase as  $r$  decreases (22, 27). This expected property is demonstrated in Fig. 1B by the straight line, which plots the theoretically predicted values of  $\text{ApEn}$  and  $\text{SampEn}$ . Second,  $\text{ApEn}$  should be independent of record length. The plot of theoretical values of  $\text{ApEn}$  and  $\text{SampEn}$  in Fig. 1B illustrates this expectation. Because of their similarity, we expect  $\text{SampEn}$  statistics to exhibit similar properties. It has been suggested that record lengths of  $10^m$ – $20^m$  should be sufficient to estimate  $\text{ApEn}(m, r)$  (27).

We tested  $\text{ApEn}$  and  $\text{SampEn}$  statistics on uniform, iid random numbers, because the results could be compared with the analytically calculated expected values. Figure 1, B and C, shows the performance of  $\text{ApEn}(2, r, N)$  and  $\text{SampEn}(2, r, N)$  on uniform iid random numbers.  $\text{SampEn}(2, r, N)$  very closely matches the expected results for  $r \geq 0.03$  and  $N \geq 100$  (Fig. 1, B and C), whereas  $\text{ApEn}(2, r, N)$  differs markedly from expectations for  $N < 1,000$  and  $r < 0.2$ . Figure 2 shows  $\text{SampEn}$  and  $\text{ApEn}$  as functions of  $r$  ( $m = 2$ ) for three sets of uniform random numbers consisting of 100, 5,000, and 20,000 points.  $\text{SampEn}$  statistics for  $r = 0.2$  are in agreement with theory for much shorter data sets. We investigated the general applicability of  $\text{SampEn}$  and  $\text{ApEn}$  statistics to random numbers with other distributions. The analysis of numbers with Gaussian, exponential, and  $\gamma$ -distributions with parameter  $\lambda = 1, 2, \dots, 10$  gave results essentially identical to those shown in Fig. 1.

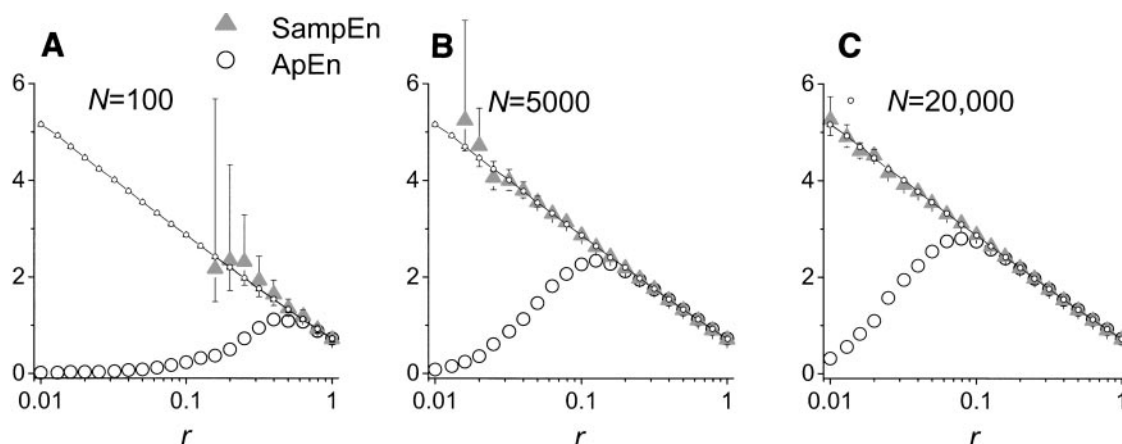


Fig. 2. Effect of record length on  $\text{SampEn}$  and  $\text{ApEn}$  as a function of  $r$  with  $m = 2$  for uniform random numbers. Confidence intervals of  $\text{SampEn}$  calculations are displayed as error bars. Plots are  $\text{SampEn}$  and  $\text{ApEn}$  as functions of  $r$  for sets of 100 (A), 5,000 (B), and 20,000 (C) points. Straight lines, theoretically predicted values of  $\text{ApEn}$  and  $\text{SampEn}$  statistics.



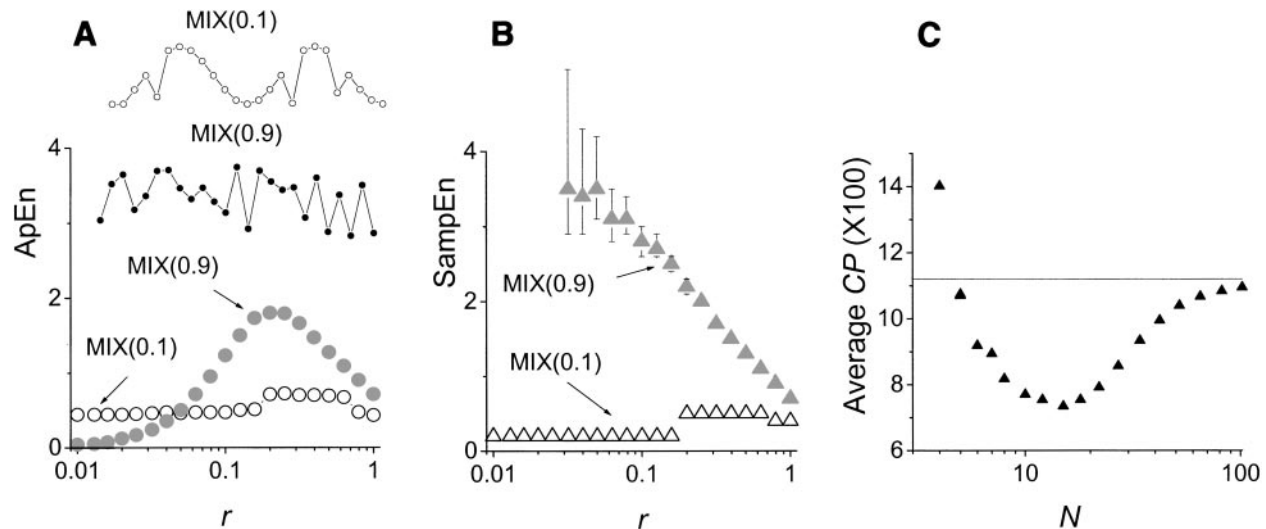


Fig. 3. Improved relative consistency but residual bias in SampEn. A and B: relative consistency is maintained by SampEn but not by ApEn for MIX(0.1) and MIX(0.9) processes. A: ApEn statistics as a function of  $r$  ( $m = 2$ ,  $N = 1000$ ) for MIX(0.1) and MIX(0.9). Inset: 25 sample points of MIX(0.1) (top) and MIX(0.9) (bottom). B: SampEn statistics for MIX(0.1) and MIX(0.9). C: residual bias of SampEn(2,0.2, $N$ ) statistics for short record lengths of independent identically distributed Gaussian numbers. For  $N = 4$ –102, SampEn(2,0.2, $N$ ) was calculated for  $\geq 10^5$  sets of random numbers and averaged.

*SampEn shows relative consistency where ApEn does not.* A critically important expected feature of ApEn is relative consistency (22, 27). That is, it is expected that, for most processes, if  $\text{ApEn}(m_1, r_1)(S) \leq \text{ApEn}(m_1, r_1)(T)$ , then  $\text{ApEn}(m_2, r_2)(S) \leq \text{ApEn}(m_2, r_2)(T)$ . That is, if record  $S$  exhibits more regularity than record  $T$  for one pair of parameters  $m$  and  $r$ , it is expected to do so for all other pairs. Graphically, plots of ApEn as a function of  $r$  for different data sets should not cross over one another. The determination that one set of data exhibits greater regularity than another can be made only when this condition is met. We tested this expectation using 1,000-point realizations of the MIX( $P$ ) process, where the degree of order could be specified. Figure 3A shows that the ApEn statistics of the less-ordered MIX(0.9), which has, on average, few matches of length  $m$  for a given template (small  $B_i$ ) for small  $r$ , rises as a function of  $r$  and crosses over a plot of ApEn statistics of the more-ordered MIX(0.1). For  $r < 0.05$ , one would conclude incorrectly that MIX(0.9) was more ordered than MIX(0.1). Thus relative consistency does not hold for ApEn statistics.

We investigated the mechanism responsible for this lack of relative consistency. Note that, for  $r = 0.5$ , ApEn statistics correctly distinguished MIX(0.1) and MIX(0.9). For this value of  $r$ , the MIX(0.1) data yielded an average of  $>46$  matches per template and  $\sim 28$  forward matches per template, whereas the MIX(0.9) data yielded  $\sim 37$  and 10, respectively. These large numbers of matches render the bias insignificant; that is, the unbiased  $A_i/B_i$  (28/46 and 10/37) is not very different from the biased  $(A_i + 1)/(B_i + 1)$  (29/47 and 11/38). As  $r$  is decreased, however, the number of template matches decreased more for the MIX(0.9) data than for the MIX(0.1) data. This made the bias significant for MIX(0.9) data under conditions for which it was insignificant in the MIX(0.1) data. For example, when  $r =$

0.05,  $\text{ApEn}(2, r, 1,000)$  of MIX(0.1) was 0.463, whereas the value for MIX(0.9) was 0.505. The spurious similarity of the ApEn statistics is due to bias. The MIX(0.1) data yielded  $\sim 27$  matches per template and 22 forward matches, whereas the MIX(0.9) data yielded only 0.45 and 0.01, respectively. Thus more than one-half of the templates of the MIX(0.9) data matched no other templates and were assigned a conditional probability of 1. In this example, ApEn statistics lack relative consistency, because less-ordered data sets have fewer matching templates and are more vulnerable to the bias generated by self-matches. SampEn analysis of the same data, on the other hand, reports correctly over the whole range of  $r$  (Fig. 3B).

*Are SampEn statistics relatively consistent?* We have shown that SampEn statistics appear to be relatively consistent over the family of MIX( $P$ ) processes, whereas ApEn statistics are not. Although we believe that relative consistency should be preserved for processes for which probabilistic character is understood, we see no general reason why ApEn or SampEn statistics should remain relatively consistent for all time series and all choices of parameters.

We propose a general, but by no means exhaustive, explanation for this phenomenon. SampEn is, in essence, an event-counting statistic, where the events are instances of vectors being similar to one another. When these events are sparse, the statistics are expected to be unstable, which might lead to a lack of relative consistency. Recall that  $\text{SampEn}(m, r, N)$  is less than or equal to  $\ln(B)$ , the natural logarithm of the number of template matches. Suppose  $\text{SampEn}(m, r, N)(S) < \text{SampEn}(m, r, N)(T)$  and that the number of  $T$ 's template matches,  $B_T$ , is less than the number of  $S$ 's template matches,  $B_S$ , which would be consistent with  $T$  displaying less order than  $S$ . Provided that  $A_T$  and  $A_S$ , the number of forward matches, are relatively large, both

SampEn statistics will be considerably lower than their upper bounds. As  $r$  decreases,  $B_T$  and  $A_T$  are expected to decrease more rapidly than  $B_S$  and  $A_S$ . Thus, as  $B_T$  becomes very small,  $\text{SampEn}(m, r, N)(T)$  will begin to decrease, approaching the value  $\ln(B_T)$ , and could cross over a graph of  $\text{SampEn}(m, r, N)(S)$ , where or while  $B_S$  is still relatively large. Furthermore, as the number of template matches decreases, small changes in the number of forward matches can have a large effect on the observed conditional probability. Thus the discrete nature of the SampEn probability estimation could lead to small degrees of crossover and intermittent failure of relative consistency, and we cannot say that SampEn will always be relatively consistent. We have shown, however, that SampEn is relatively consistent for conditions where ApEn is not, and we have not observed any circumstance where ApEn maintains relative consistency and SampEn does not.

*One source of the residual bias in SampEn is correlation of templates.* Although more consonant with theory than ApEn, we found that SampEn statistics deviated from predictions for very short data sets. For  $10^5$  sets of Gaussian random numbers with  $m = 2$ , and  $r = 0.2$ , we found that the deviation was  $<3\%$  for record lengths  $>100$  but as high as  $35\%$  for sets of 15 points. Figure 3C shows the biased results of  $\text{SampEn}(2, 0.2, N)$  for the range of  $4 \leq N \leq 100$ . We suspected that the integral expressions for the parameters  $\text{ApEn}(m, r)$  and  $\text{SampEn}(m, r)$  could not be used as expected values of the statistics  $\text{ApEn}(m, r, N)$  and  $\text{SampEn}(m, r, N)$  under all conditions, because the expressions relied on the assumption that the templates were independent of one another. As  $N$  decreases and  $m$  increases, however, a larger proportion of templates are comprised of overlapping segments of the record and are thus not independent. Because of this correlation, results might deviate from these predictions for short data sets. We thus tested the hypothesis that the majority of the bias results from nonindependence of the templates.

One way to test the hypothesis is to compare observed values of SampEn with those obtained from a model accounting for template correlation. The hypothesis predicts that the values should match. We tested the simplest case of  $m = 2$  and  $N = 4$ , where a data set can be represented by  $\{w, x, y, z\}$ , so that there are exactly two vectors of length  $m = 3$ ,  $(w, x, y)$  and  $(x, y, z)$ . If  $(w, x)$  is close to  $(x, y)$ , it stands to reason that  $(w, x, y)$  will have a higher than expected probability of being close to  $(x, y, z)$ . Formally, the conditional probability that  $(w, x, y)$  and  $(x, y, z)$  will be close given  $(w, x)$  and  $(x, y)$  are close is  $[\int_{-\infty}^{\infty} p(w) \int_{w-rSD}^{w+rSD} p(x) \int_{x-rSD}^{x+rSD} p(y) \int_{y-rSD}^{y+rSD} p(z) dz dy dx dw] / [\int_{-\infty}^{\infty} p(w) \int_{w-rSD}^{w+rSD} p(x) \int_{x-rSD}^{x+rSD} p(y) dy dx dw]$ , where  $p(z)$  is the Gaussian probability density function. For  $r = 0.2$ , this was evaluated by numerical integration and found to be 0.137. For  $10^6$  sets of 4 iid Gaussian numbers, we calculated  $\text{SampEn}(2, 0.2, 4)$  to be 0.140. Thus the observed and expected results nearly agree, indicating that bias due to nonindependence of templates accounts for most of the observed deviation from the conditional probability

of 0.112 expected for independent templates in this case.

A second way to test the hypothesis is to calculate SampEn statistics for one set of data under two conditions: with and without overlapping templates. The hypothesis predicts that the results should not match. For this test, we chose the case of  $m = 2$  and  $N = 6$ , the shortest record containing two nonoverlapping templates of length  $m + 1 = 3$ . We can represent each set as  $\{a, b, c, x, y, z\}$ . For  $10^6$  sets of six Gaussian random numbers, we calculated the conditional probability that  $(a, b, c)$  was within  $r$  of  $(x, y, z)$  given that their first two points were close, thus calculating the probability for pairs of disjoint templates. The result was 0.111, very close to the expected 0.112 for independent templates. For the same number sets, we then calculated the average value of  $\text{SampEn}(2, 0.2, 6)$ , and the result was 0.094. Thus the two results do not match, in support of the hypothesis. We conclude from this analysis that the statistics  $\text{SampEn}(m, r, N)$  are not completely unbiased under all conditions and that the bias of SampEn for very small data sets is largely due to nonindependence of templates.

One method for removing this bias would be to partition the time series  $\{u_j | 1 \leq j \leq N\}$  into the  $m + 1$  sets of neighboring, disjoint vectors of length  $m + 1$   $\mathbf{X}_i = [\mathbf{u}_{i+k(m+1)}, \mathbf{u}_{i+k(m+1)+1}, \dots, \mathbf{u}_{i+k(m+1)+m}]$ ;  $0 \leq k \leq [N - (m + 1)] / (m + 1)$ , where  $i$ , the initial point of the first template, ranges from 1 to  $m + 1$ . The conditional probability that vectors close for  $m$  points remain close at the next point would be calculated for each of the sets of vectors  $\mathbf{X}_i$  and then averaged. Because this calculation compares only disjoint templates, it will not suffer from the bias introduced by nonindependent templates. This truly unbiased approach has the potentially severe limitation of reducing the number of possible template matches and enlarging the confidence intervals about the SampEn estimate. Because this bias appears to be present only for very small  $N$ , the disjoint template approach does not appear necessary in usual practice.

*Cross-SampEn shows relative consistency where cross-ApEn does not.* As noted above, an essential feature of the measures of order is their relative consistency. That is, if one series is more ordered than another, it should have lower values of ApEn and SampEn for all conditions. We can extend this idea to cross-ApEn and cross-SampEn; if a pair of series is more synchronous than another pair, it should have lower values of cross-ApEn and cross-SampEn statistics for all conditions tested.

We tested the ability of cross-ApEn and cross-SampEn to distinguish between MIX(0.1) and the less-ordered MIX(0.6) processes. The strategy was to compare each with the intermediate MIX(0.3) process. The expected result was that the [MIX(0.3), MIX(0.1)] pair should appear more ordered than the [MIX(0.3), MIX(0.6)] pair, because MIX(0.1) is significantly more ordered than MIX(0.6). That is,  $\text{cross-ApEn}(2, r, 250) [\text{MIX}(0.3) \parallel \text{MIX}(0.1)]$  should be less than  $\text{cross-ApEn}(2, r, 250) [\text{MIX}(0.3) \parallel \text{MIX}(0.6)]$ , and cross-



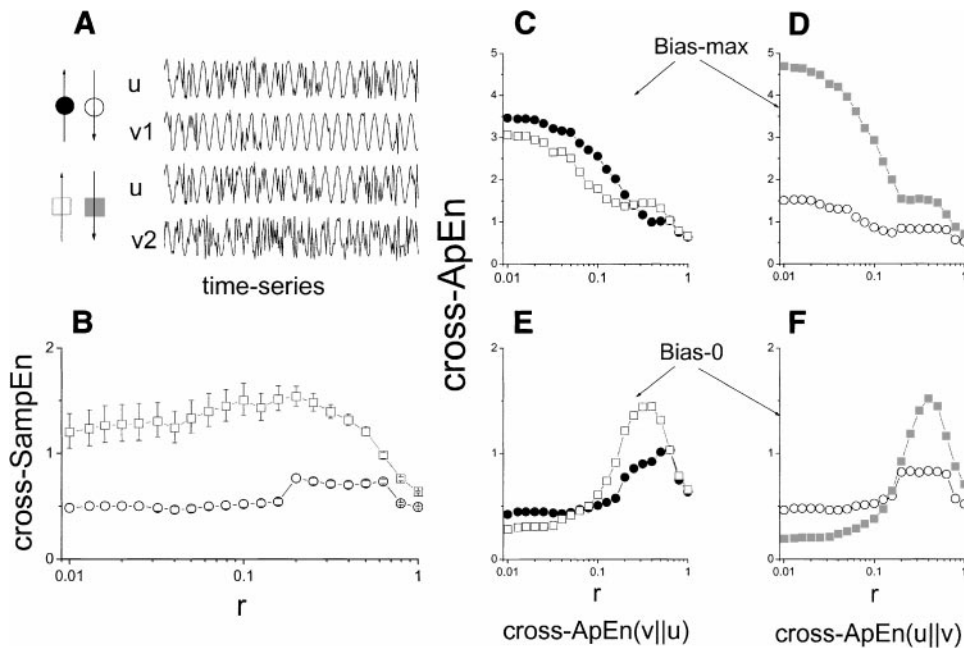


Fig. 4. Results of cross-ApEn analysis of MIX( $P$ ) processes are dependent on direction of analysis and on correction strategy. **A:** key showing pairs of time series evaluated, linked by an arrow indicating direction of analysis (arrow points from target series to template series). **B:** cross-SampEn(1, $r$ ,250)[MIX(0.3)||MIX(0.1)] and cross-SampEn(1, $r$ ,250)[MIX(0.3)||MIX(0.6)]. No correction strategy is required. **C:** cross-ApEn(1, $r$ ,250) [MIX(0.3)||MIX(0.1)] (●) and cross-ApEn(1, $r$ ,250) [MIX(0.3)||MIX(0.6)] (□) with bias-max correction. **D:** identical to C, but with direction of analysis reversed, showing cross-ApEn(1, $r$ ,250)[MIX(0.1)||MIX(0.3)] instead of cross-ApEn(1, $r$ ,250)[MIX(0.3)||MIX(0.1)]. **E:** identical to C, but with bias-0, rather than bias-max, correction. **F:** identical to D, but with bias-0, rather than bias-max, correction.

SampEn(2,  $r$ , 250) [MIX(0.3)||MIX(0.1)] should be less than cross-SampEn(2,  $r$ , 250) [MIX(0.3)||MIX(0.6)].

We tested this prediction bidirectionally, that is, MIX(0.3) served as the template series for one analysis and as the target series for the next, and over a range of tolerances  $r$  by using the bias-max and bias-0 strategies for ensuring that cross-ApEn was defined. The results are shown in Fig. 4. MIX(0.3) was the template series in Fig. 4, C and E, and the target series in Fig. 4, D and F.

The expected result is that cross-ApEn and cross-SampEn should be less for the [MIX(0.3), MIX(0.1)] pair than for the [MIX(0.3), MIX(0.6)] pair. That is, the circles should always be below the squares. We found this to be true for only one of the four tests of cross-ApEn, the case of using MIX(0.3) as the target series with the bias-max correction strategy (Fig. 4D). In the

other cases, the order of results was reversed (Fig. 4C) or crossed over (Fig. 4, E and F). As shown in Fig. 4B, cross-SampEn returned the expected results with a high degree of confidence across the range of tolerances  $r$ .

Thus cross-ApEn statistics fail as a means of judging the relative order of two time series by their similarity to a third series. In practice, however, cross-ApEn has been used differently: to determine the relative synchrony of two pairs of time series of clinical data from different patients. We thus tested cross-ApEn and cross-SampEn on two sections of a long multivariate cardiovascular time series used in the 1991 Santa Fe competition for time series forecasting (35). The series consisted of concurrent measurements of a sleeping patient's heart rate (hr) and chest volume (cv). We

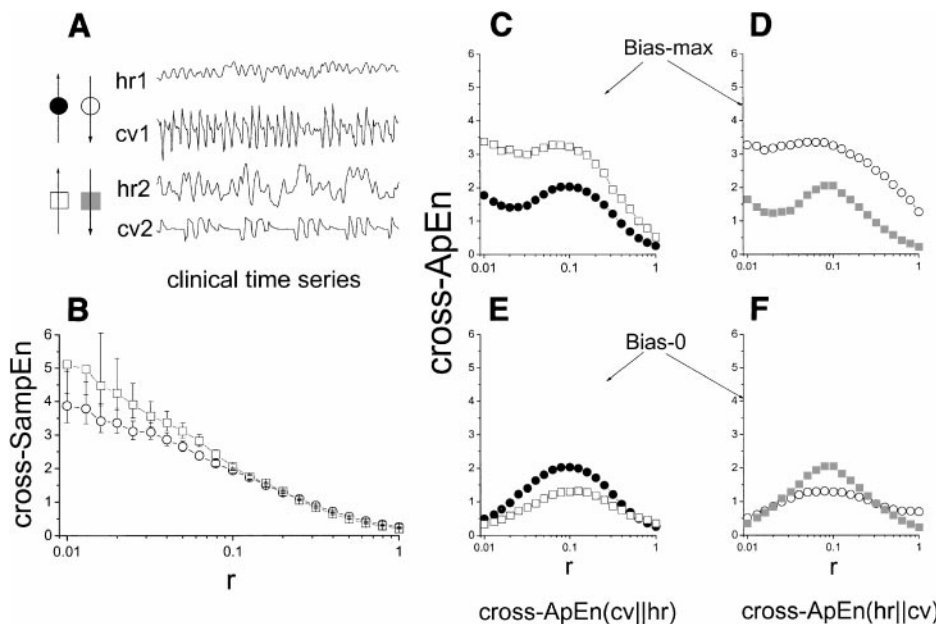


Fig. 5. Cross-ApEn results can vary greatly because of direction effects and correction strategy employed. Shown are calculations of cross-ApEn statistics of two 250-point segments of multivariate clinical cardiovascular time series used in 1991 Santa Fe time-series forecasting competition. Cross-ApEn was calculated for concurrent series of heart rate (hr) and chest volume (cv). **A:** key to time series and direction of analysis (same conventions as Fig. 4A). **B:** cross-SampEn(1, $r$ ,250)(cv||hr) of 2 pairs of series. **C:** cross-ApEn(1, $r$ ,250)(cv||hr) for 2 segments of time series with bias-max correction. **D:** same as C, but with direction of analysis reversed to show cross-ApEn(1, $r$ ,250)(hr||cv), rather than cross-ApEn(1, $r$ ,250)(cv||hr). **E:** identical to C, but with bias-0, rather than bias-max, correction. **F:** identical to D, but with bias-0, instead of bias-max, correction. Distinctions made by cross-ApEn are large but highly dependent on correction method employed and on direction of analysis. Cross-SampEn distinguishes these clinical data in a consistent fashion.

compared the pairs (hr1,cv1) and (hr2,cv2) shown in Fig. 5A, excerpted from the larger time series. Here, the expected result was not known beforehand, and our question was whether one pair consistently appeared more synchronous than the other. This is an extension of the expected relative consistency of ApEn discussed above.

Figure 5 shows the results for  $N = 250$ ,  $m = 1$ , and a range of  $r$ . We set  $m = 1$ , in accordance with published practice for analyzing series of similar length (28), and the record length of  $N = 250$  exceeds the  $10^m$ – $20^m$  points recommended for ApEn analysis (27). Time series are shown in Fig. 5A. The question is whether the pair (hr1,cv1) has more joint synchrony than the pair (hr2,cv2). The expected result is that the circles should be consistently higher or lower than the squares for Fig. 5, D–F. This was not the case; conclusions about relative synchrony by use of cross-ApEn analysis depended on which series served as the template and on the correction strategy, and there was no consistent result. Cross-SampEn, on the other hand, consistently reported that (hr1,cv1) had more joint synchrony than (hr2,cv2) (Fig. 5B).

For 32 sets of these data, we further examined the correction methods, calculating cross-ApEn( $m, r, N$ ) (cv||hr) and cross-ApEn( $m, r, N$ ) (hr||cv) for each set. For the relaxed condition of  $r = 1.0$ , we found that cross-ApEn(1,1,250)(cv||hr) was defined for only 19 of the 32 cases, whereas cross-ApEn(1,1,250)(hr||cv) was defined for only 12 cases. For the more stringent case of  $r = 0.2$ , cross-ApEn(1,0.2,250)(cv||hr) was never defined, whereas cross-ApEn(1,0.2,250) (hr||cv) was defined for 2 of the 32 cases. By contrast, for all  $r \geq 0.08$ , cross-SampEn(1, $r$ ,250)(cv||hr) was defined for each of the 32 pairs. Thus cross-SampEn had a more consistent performance for evaluating these clinical cardiovascular data.

**Summary.** We have developed and characterized SampEn, a new family of statistics measuring complexity and regularity of clinical and experimental time series data and compared it with ApEn, a similar family. We find that SampEn statistics 1) agree much better than ApEn statistics with theory for random numbers with known probabilistic character over a broad range of operating conditions, 2) maintain relative consistency where ApEn statistics do not, and 3) have residual bias for very short record lengths, in a large part because of nonindependence of templates. Furthermore, cross-SampEn is a more consistent measure of joint synchrony of pairs of clinical cardiovascular time series. We attribute the difficulties of ApEn analysis to the practice of counting self-matches and of cross-ApEn to the problem of unmatched templates resulting in undefined probabilities. The differences are that SampEn does not count templates as matching themselves and does not employ a template-wise strategy for calculating probabilities. SampEn statistics provide an improved evaluation of time series regularity and should be a useful tool in studies of the dynamics of human cardiovascular physiology.

We thank L. Pitt, D. Scollan, and Rizwan-uddin for advice and Virginia's Center for Innovative Technology for support.

Address for reprint requests and other correspondence: J. R. Moorman, Box 6012, MR4 Bldg., UVAHSC, Charlottesville, VA 22908 (E-mail: rmoorman@virginia.edu).

Received 2 August 1999; accepted in final form 13 December 1999.

## REFERENCES

1. **Bendat JS and Piersol AG.** *Random Data. Analysis and Measurement Procedures.* New York: Wiley, 1986.
2. **Ben-Mizrachi A, Procaccia I, and Grassberger P.** The characterization of experimental (noisy) strange attractors. *Phys Rev A* 29A: 975, 1984.
3. **Dawes GS, Moulden M, Sheil O, and Redman CW.** Approximate entropy, a statistic of regularity, applied to fetal heart rate data before and during labor. *Obstet Gynecol* 80: 763–768, 1992.
4. **Durrett R.** *Probability Theories and Examples.* Belmont, CA: Duxbury, 1996, p. 14.
5. **Eckmann JP and Ruelle D.** Ergodic theory of chaos and strange attractors. *Rev Modern Phys* 57: 617–654, 1985.
6. **Fleisher LA, DiPietro JA, Johnson TR, and Pincus S.** Complementary and noncoincident increases in heart rate variability and irregularity during fetal development. *Clin Sci* 92: 345–349, 1997.
7. **Fleisher LA, Pincus SM, and Rosenbaum SH.** Approximate entropy of heart rate as a correlate of postoperative ventricular dysfunction. *Anesthesiology* 78: 683–692, 1993.
8. **Goldberger AL, Mietus JE, Rigney DR, Wood ML, and Fortney SM.** Effects of head-down bed rest on complex heart rate variability: response to LBNP testing. *J Appl Physiol* 77: 2863–2869, 1994.
9. **Grassberger P.** Finite sample corrections to entropy and dimension estimates. *Phys Lett A* 128: 369, 1988.
10. **Grassberger P and Procaccia I.** Estimation of the Kolmogorov entropy from a chaotic signal. *Phys Rev A* 28: 2591–2593, 1983.
11. **Grassberger P, Schreiber T, and Schaffrath C.** Nonlinear time sequence analysis. *Int J Bifur Chaos* 1: 547, 1991.
12. **Ho KK, Moody GB, Peng CK, Mietus JE, Larson MG, Levy D, and Goldberger AL.** Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. *Circulation* 96: 842–848, 1997.
13. **Hogue CJ, Domitrovich PP, Stein PK, Despotis GD, Re L, Schuessler RB, Kleiger RE, and Rottman JN.** RR interval dynamics before atrial fibrillation in patients after coronary artery bypass graft surgery. *Circulation* 98: 429–434, 1998.
14. **Korpelainen JT, Sotaniemi KA, Makikallio A, Huikuri HV, and Myllyla VV.** Dynamic behavior of heart rate in ischemic stroke. *Stroke* 30: 1008–1013, 1999.
15. **Lipsitz LA, Pincus SM, Morin RJ, Tong S, Eberle LP, and Gootman PM.** Preliminary evidence for the evolution in complexity of heart rate dynamics during autonomic maturation in neonatal swine. *J Auton Nerv Syst* 65: 1–9, 1997.
16. **Makikallio TH, Ristimäe T, Airaksinen KE, Peng CK, Goldberger AL, and Huikuri HV.** Heart rate dynamics in patients with stable angina pectoris and utility of fractal and complexity measures. *Am J Cardiol* 81: 27–31, 1998.
17. **Makikallio TH, Seppanen T, Niemela M, Airaksinen KE, Tulppo M, and Huikuri HV.** Abnormalities in beat-to-beat complexity of heart rate dynamics in patients with a previous myocardial infarction. *J Am Coll Cardiol* 28: 1005–1011, 1996.
18. **Nelson JC, Rizwan-uddin, Griffin MP, and Moorman JR.** Probing the order within neonatal heart rate variability. *Pediatr Res* 43: 823–831, 1998.
19. **Palazzolo JA, Estafanous FG, and Murray PA.** Entropy measures of heart rate variation in conscious dogs. *Am J Physiol Heart Circ Physiol* 274: H1099–H1105, 1998.
20. **Pincus S and Singer BH.** Randomness and degrees of irregularity. *Proc Natl Acad Sci USA* 93: 2083–2088, 1995.
21. **Pincus SM.** Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 88: 2297–2301, 1991.
22. **Pincus SM.** Approximate entropy (ApEn) as a complexity measure. *Chaos* 5: 110–117, 1995.

23. **Pincus SM.** Quantifying complexity and regularity of neurobiological systems. *Methods Neurosci* 28: 336–363, 1995.
24. **Pincus SM, Cummins TR, and Haddad GG.** Heart rate control in normal and aborted-SIDS infants. *Am J Physiol Regulatory Integrative Comp Physiol* 264: R638–R646, 1993.
26. **Pincus SM, Gladstone IM, and Ehrenkranz RA.** A regularity statistic for medical data analysis. *J Clin Monit* 7: 335–345, 1991.
27. **Pincus SM, and Goldberger AL.** Physiological time-series analysis: what does regularity quantify? *Am J Physiol Heart Circ Physiol* 266: H1643–H1656, 1994.
28. **Pincus SM, Mulligan T, Iranmanesh A, Gheorghiu S, Godschalk M, and Veldhuis JD.** Older males secrete luteinizing hormone and testosterone more irregularly, and jointly more asynchronously, than younger males. *Proc Natl Acad Sci USA* 93: 14100–14105, 1996.
29. **Pincus SM and Viscarello RR.** Approximate entropy: a regularity measure for fetal heart rate analysis. *Obstet Gynecol* 79: 249–255, 1992.
30. **Press WH, Teukolsky SA, Vetterling WT, and Flannery BP.** *Numerical Recipes in FORTRAN. The Art of Scientific Computing.* New York: Cambridge University Press, 1994.
31. **Rosner B.** *Fundamentals of Biostatistics.* Boston, MA: PWS-Kent, 1990, p. 161.
32. **Ryan SM, Goldberger AL, Pincus SM, Mietus J, and Lipsitz LA.** Gender- and age-related differences in heart rate dynamics: are women more complex than men? *J Am Coll Cardiol* 24: 1700–1707, 1994.
33. **Schuckers SA.** Use of approximate entropy measurements to classify ventricular tachycardia and fibrillation. *J Electrocardiol* 31 Suppl: 101–105, 1998.
34. **Tulppo MP, Makikallio TH, Takala TE, Seppanen T, and Huikuri HV.** Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *Am J Physiol Heart Circ Physiol* 271: H244–H252, 1996.
35. **Weigend AS and Gershenfeld NA.** *Time Series Prediction: Forecasting the Future and Understanding the Past.* Boston, MA: Addison-Wesley, 1994.

